

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
22 November 2001 (22.11.2001)

PCT

(10) International Publication Number
WO 01/88114 A2

(51) International Patent Classification²: **C12N 15/00**

LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX,
MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL,
TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(21) International Application Number: **PCT/EP01/05794**

(84) Designated States (*regional*): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,
CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(22) International Filing Date: 18 May 2001 (18.05.2001)

Declarations under Rule 4.17:

- as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(i)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG)
- of inventorship (Rule 4.17(iv)) for US only

(25) Filing Language: English

Published:

- without international search report and to be republished upon receipt of that report

(26) Publication Language: English

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(30) Priority Data:
0012231.7 19 May 2000 (19.05.2000) GB

(71) Applicant (for all designated States except US): DEV-GEN NV [BE/BE]; Technologiepark 9, B-9052 Zwijnaarde (BE).

(72) Inventors; and

(75) Inventors/Applicants (for US only): PLAETINCK, Geert [BE/BE]; Pontstraat 16, B-9820 Merelbeke (BE). RENARD, Jean-Pierre [BE/BE]; Peter Benoitaan 141, B-9050 Gentbrugge (BE). MORTIER, Katherine [BE/BE]; Lange Boomgaardstraat 25, B- Gent (BE). BOGAERT, Thierry [BE/BE]; Wolvendreef 26g, B-8500 Kortrijk (BE).

(74) Agent: BOULT WADE TENNANT; Verulam Gardens, 70 Gray's Inn Road, London WC1X 8BT (GB).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK,

A2

WO 01/88114 A2

(54) Title: GENE EXPRESSION SYSTEM

(57) Abstract: There are provided DNA constructs, including replicable cloning vectors and expression vectors, comprising a bacteriophage promoter operably linked to an outtron sequence. The expression vectors provided by the invention are useful in the expression of recombinant polypeptides in host cells or organisms and are particularly useful in expression of recombinant polypeptides in nematode worms such as *C. elegans*.

NUC 38884

- 1 -

GENE EXPRESSION SYSTEM

Field of the invention

5 The invention relates to the expression of DNA, genes, cDNAs, proteins, peptides and parts thereof in the nematode worm *C. elegans*. In particular, the invention relates to methods of improving the translation of RNAs transcribed in *C. elegans* using a
10 bacteriophage polymerase by introduction of a trans-splice recognition site recognised by an SL1 trans-splice recognition sequence into the DNA template transcribed by the bacteriophage polymerase.

15 Background to the invention

Eukaryotic versus prokaryotic expression.

20 Bacteriophage RNA polymerases, such as T7, T3, and SP6, and their corresponding promoters have been used extensively to drive the expression of heterologous genes in a variety of organisms. In co-pending International patent application No. WO 00/01846, Plaetinck et al. describe the use of the T7 system to express DNA, genes, cDNA, proteins and peptides of parts thereof and for the expression of double-stranded RNA (dsRNA) in the nematode model system *C. elegans*.

25 The bacteriophage expression systems are well known in the art for use in prokaryotic host cells, such as *E. coli*, and have the advantage that they provide simple and strong expression systems dependent only on one RNA polymerase and one well defined promoter. The application of such efficient expression systems in eukaryotic organisms is, however, not evident, mainly because messenger RNAs from eukaryotes and prokaryotes have a different

- 2 -

structure, which has implications for translation efficiency and RNA stability.

Messenger RNAs of higher eukaryotes share a functionally essential 5' CAP structure. This
5 structure is generated during a capping reaction that is linked exclusively to RNA polymerase II transcription. Prokaryotic RNA polymerases such as bacteriophage T3, T7 and SP6 polymerases do not provide messenger RNAs with such a CAP structure,
10 leading to inefficient translation in eukaryotic systems (Fuerst et al. J. Mol. Biol:206:333-348 (1989)).

One way to improve translation of uncapped mRNAs in eukaryotic systems is by the insertion of an
15 internal ribosome entry site (IRES) sequence 5' of the coding sequence. For example, Elroy-Stein, et al., Proc. Natl. Acad. Sci. USA 87:6743-6747 (1990), describe the cloning of the untranslated region of the ECMV virus downstream of the T7 promoter in order to
20 enhance the efficiency of translation. In other systems translation of T7-derived transcripts may be enhanced by addition of a CAP structure derived from a capped transcript. For example, in Trypanosoma a 5' CAP structure is added to T7 generated RNA transcripts
25 by a natural occurring trans-splicing reaction (Wirtz et al. NAR 22:3887-3894 (1994)).

Trans-splicing in C. elegans.

In C. elegans many mRNAs contain an identical
30 short leader sequence, designated the spliced leader (SL). This splice leader is donated by a small RNA (SL RNA) via a trans-splicing reaction. This trans
splicing was first observed by Krause et al., Cell 49:753-61 (1987). The splice leader RNA exists as a
35 small nuclear ribonucleoprotein particle and has the trimethylguanosine cap that is characteristic of

- 3 -

eukaryotic small nuclear RNAs. The trimethylguanosine cap present on the spliced leader RNA is transferred to the pre-mRNA during the trans-splicing reaction. Thereafter, the trimethylguanosine cap is maintained 5 on the mature mRNA (Van Doren et al., Mol. Cell. Biol. 10:1769-1772 (1990). The trans-splicing signal for such a splice leader is essentially an intron missing only the 5' splice site, designated an 'outtron'. An outtron has essentially all the intron sequence 10 including a trans-splice acceptor site homologous to a UUUCAG sequence preceded by a AU rich region (Conrad et al., NAR 21:913-919 (1993)). Introduction of an outtron into the 5' untranslated region of a *C. elegans* gene converts it to a trans-spliced gene (Conrad et 15 al., EMBO J. 12:1249-1255 (1993); Conrad et al. Mol. Cell Biol. 11:1931-1926 (1991)) and introduction of donor sites in a natural trans-spliced *C. elegans* gene prevents trans-splicing and converts it into a more conventional gene.

20

Description of the invention.

Until recently, expression of heterologous and homologous genes in *C. elegans* was mainly achieved by linking an appropriate coding sequence to a selected 25 *C. elegans* promoter. The present inventors have recently demonstrated that the recombinant gene expression in *C. elegans* can be based on the prokaryotic T7 expression system (WO 00/01846). However, the present inventors found that the 30 expression system was far from being efficient, or at least the resulting expression was much lower than would be expected from this T7 related expression system. It was concluded that this low expression was mainly due to RNA instability or translation arrest. Furthermore, it was reasoned that fundamental 35 differences between prokaryotic and eukaryotic

- 4 -

expression systems, particularly the requirement for capping of the 5' end of the mRNA for efficient translation in eukaryotic systems, was the main reason for this unexpectedly low expression.

5 The inventors have now developed a solution to the problem of the inefficiency of the T7 system in eukaryotic host cells and organisms, particularly in *C. elegans*, and have constructed a generally applicable expression system which allows for the
10 efficient expression of genes, DNA, cDNA, peptides and proteins under the regulation of the T7 promoter in *C. elegans*.

Therefore, in accordance with a first aspect of the invention there is provided a DNA construct
15 comprising a bacteriophage promoter operably linked to an outron sequence.

It is an essential feature of the DNA construct of the invention that the bacteriophage promoter and the outron sequence are "operably linked", that is to say they are arranged in a relationship permitting them to function in their intended manner. In this case, the bacteriophage promoter is positioned upstream of the outron sequence such that it is capable of promoting transcription of the outron sequence upon binding of an appropriate RNA polymerase, with the outron sequence forming the extreme 5' end of the resulting transcript.

The DNA construct may further comprise at least one restriction enzyme recognition site positioned downstream of and proximal to the outron sequence.
30 Advantageously, the DNA construct may contain multiple restriction sites forming a multi-cloning site. The purpose of the restriction site/multi-cloning site is to facilitate cloning of a heterologous or homologous DNA fragment downstream of the outron sequence. A DNA construct comprising a bacteriophage promoter, an outron sequence and a restriction site/multi-cloning

- 5 -

site may therefore be referred to hereinafter as an 'outron cloning construct'.

In an outron cloning construct it is advantageous for the restriction site/multi-cloning site to be positioned fairly proximal to the outron sequence (e.g. within 100bp) such that a heterologous or homologous sequence inserted at this site may be co-transcribed with the outron sequence on a single mRNA. However, further sequence elements may be interposed between the outron sequence and the restriction site/multi-cloning site. For example, the general purpose vector pDW3123 described in the accompanying examples has a synthetic intron A sequence between the outron sequence and the multi-cloning site.

In one preferred embodiment of the invention, the DNA construct is a replicable cloning vector, such as, for example, a plasmid vector. In addition to the bacteriophage promoter, outron sequence and optional restriction site/multi-cloning site, the vector may further contain one or more of the general features commonly found in cloning vectors, for example an origin of replication to allow autonomous replication within a host cell and a selective marker, such as an antibiotic resistance gene. Although not essential, the vector may also contain a poly-adenylation signal to stabilize and process the 3' end of the mRNA transcribed from the bacteriophage promoter. A preferred example is the 3'UTR from the *C. elegans* unc-54 gene, but any other 3'UTR or polyadenylation signal may be used.

Outron-containing DNA constructs according to the invention may be easily be constructed from the component sequence elements using standard recombinant techniques well known in the art and described, for example, in F. M. Ausubel et al. (eds.), *Current Protocols in Molecular Biology*, John Wiley & Sons, Inc. (1994).

- 6 -

Outron sequences for use in the constructs of the invention may be isolated from natural *C. elegans* genes using standard molecular biology techniques. For example, a natural outtron sequence might be 5 amplified using the polymerase chain reaction or an equivalent amplification technique using *C. elegans* genomic DNA as a template. Alternatively, synthetic outtron sequences may be synthesised, for example, by annealing two complementary single 10 stranded oligonucleotides, as illustrated in the accompanying examples. Once a DNA fragment comprising the outtron sequence has been obtained it would be a matter of routine to assemble an outtron construct by linking the outtron in the correct orientation relative 15 to the bacteriophage promoter.

The sequences of the commonly used bacteriophage promoters, e.g. T7, T3 and SP6, are well known in the art and oligonucleotides containing functional phage promoter sequences can be readily synthesised using 20 standard oligonucleotide synthesis techniques. It would be a matter of routine to insert such a synthetic promoter sequence into, for example, a plasmid vector backbone containing, for example, an origin of replication a selective marker and a 25 suitable restriction site. Alternatively, one of the many plasmid vectors containing bacteriophage promoter sequences known in the art may be used as the starting point for the construction of a plasmid-based outtron cloning vector. The known vectors generally contain, 30 in addition to the phage promoter sequence, one or more restriction sites conveniently positioned downstream of the phage promoter and also a bacterial origin of replication and a selective marker. Once the vector backbone is in place the outtron sequence 35 may simply be inserted in the appropriate position downstream of the bacteriophage promoter.

In a particularly useful embodiment the invention

- 7 -

provides a DNA construct for use in bacteriophage promoter-driven expression of a polypeptide in a eukaryotic host cell or organism. This construct comprises a bacteriophage promoter operably linked to
5 a DNA sequence such that it is capable of initiating transcription of the DNA sequence upon binding of an appropriate RNA polymerase to the promoter, wherein the aforesaid DNA sequence comprises an outron sequence and at least one open reading frame
10 positioned downstream of the outron sequence.

The open reading frame may be essentially any protein-encoding DNA sequence bounded by start and stop codons. This protein-encoding DNA sequence may include introns, as both trans-splicing and cis-splicing can occur together.
15

A DNA construct according to this embodiment of the invention, which may be referred to hereinafter as an 'outron expression construct', may be derived from an outron cloning construct by insertion of a
20 heterologous or homologous protein-encoding DNA fragment into the restriction site/multi-cloning site. It is essential that the heterologous or homologous DNA fragment be inserted downstream of the outron sequence such that the two sequences may be co-
25 transcribed, with the outron sequence forming part of the 5' untranslated region of the resulting mRNA.

The outron expression construct may advantageously form an expression vector, such as, for example, a plasmid vector.. Most preferably, the
30 expression vector will be one suitable for use in the nematode worm *C. elegans*. In addition to the bacteriophage promoter, outron sequence and protein-encoding DNA sequence (open reading frame), the expression vector may further contain one or more of
35 the general features commonly found in expression vectors, for example an origin of replication to allow autonomous replication within a bacterial host cell

- 8 -

and a selective marker, such as an antibiotic resistance gene. The vector may also contain a polyadenylation signal to stabilize and process the 3' end of the mRNA transcribed from the bacteriophage 5 promoter. A preferred example is the 3'UTR from the *C. elegans* unc-54 gene, but any other 3'UTR or polyadenylation signal may be used. An additional element, such as for example a synthetic intron, may be interposed between the outron sequence and the open 10 reading frame.

It is important that the open reading frame is positioned downstream of and proximal to the outron sequence in the expression construct such that (i) the two elements are co-transcribed to form a single mRNA 15 and (ii) the outron sequence forms part of the 5' untranslated region of the mRNA. If the appropriate splicing machinery and a supply of SL RNAs is provided by the eukaryotic host cell or organism then the uncapped 5' end of the pre-mRNA transcribed from the 20 expression construct will be replaced with a capped splice leader via the trans-splicing reaction. This will greatly increase the efficiency of translation in a eukaryotic host system.

The use of an outron sequence at the extreme 5' 25 end of the RNA provides a solution to the problem of reduced expression efficiency in eukaryotic systems wherever the type of promoter/polymerase used to drive gene expression leads to the production of uncapped transcripts, provided that the host cell or organism 30 produces the spliced leader RNAs required for the trans-splicing reaction.

Outron sequences which may be utilised in accordance with the invention include naturally occurring outron sequences isolated from SL1-specific 35 *C. elegans* genes (Conrad, R. Functional analysis of a *C. elegans* trans-splice acceptor. *Nucleic Acids Res.* 1993, 21(4), pp913-919; Conrad, R. SL1 trans-splicing

- 9 -

specified by AU-rich synthetic RNA inserted at the 5' end of *Caenorhabditis elegans* pre-mRNA. RNA. 1995, 1(2), pp164-170) and also synthetic outron sequences which are functionally equivalent to the natural *C. elegans* outron sequences, including variants of naturally occurring *C. elegans* outrons. The phrase "functionally equivalent" means that the synthetic intron is recognised by the *C. elegans* trans-splicing machinery and can be trans-spliced to a *C. elegans* splice leader RNA, preferably the SL1 splice leader.

Experimental evidence indicates that trans-splicing in *C. elegans* is signalled by an AU-rich intron-like sequence followed by a splice acceptor site (Conrad et al 1993 and 1995). For the purposes of the present application the terms "outron" or "outron sequence" should be interpreted as referring to both the AU-rich region from the 5' end of the pre-mRNA to the trans-splice acceptor site and the trans-splice acceptor site itself. In connection with the DNA constructs of the invention, the terms "outron" and "outron sequence" refer to features present in the DNA which encodes the pre-mRNA.

The consensus splice acceptor site for trans-splicing of outrons and the consensus 3' splice acceptor site for cis-splicing of introns are essentially identical (UUUCAG). Moreover, a normally trans-spliced acceptor site can be efficiently cis-spliced when a donor splice site is inserted upstream within the outron sequence. It is therefore important that the outron constructs described herein do not contain any potential splice donor sequence upstream of the splice acceptor within the outron and downstream of the transcription start site such that it will be transcribed in the mRNA encoded by the construct. If such a site were present than there would be a potential for cis-splicing rather than

- 10 -

trans-splicing.

It has also been observed that the overall length of the outron has an effect on the efficiency of trans-splicing, longer outrons in general working
5 better than shorter ones (Conrad et al. 1995).

Advantageously, the outron sequences for inclusion into the outron constructs described herein should be greater than about 50nt in length.

10 A synthetic outron containing an AT stretch and a TTTTCAG sequence has been shown to be functional in *C. elegans*. As illustrated in the accompanying Examples, the insertion of an outron sequence into the 5' untranslated region of GFP reporter construct, downstream of the promoter and upstream of the GFP
15 open reading frame, is required for optimal expression of bacteriophage RNA polymerase transcribed reporter gene mRNA in *C. elegans*.

20 Suitable bacteriophage promoters which may be used in the DNA constructs according to the invention include T7, T3 and SP6 promoters, with T7 being the most preferred. As discussed above, these bacteriophage promoters have long been known to be useful tools in molecular biology since they can provide simple and strong expression systems dependent
25 only on the binding of the specific or cognate RNA polymerase.

30 In a still further aspect, the invention provides a method for expressing a recombinant polypeptide in *C. elegans*, which method comprises:

introducing an outron expression construct, as described above, said construct being an expression vector suitable for use in *C. elegans*, into a *C. elegans* strain which expresses an RNA polymerase
35 specific for the bacteriophage promoter present in said DNA construct in one or more tissues or cell

- 11 -

types.

An outron expression vector for use in this method may be constructed by inserting DNA encoding the polypeptide of interest into an outron cloning 5 vector, as described above. The vector must be one which is suitable for use in *C. elegans*, plasmid-based vectors are the most preferred.

The *C. elegans* worms are preferably transgenic worms carrying a transgene capable of expressing the 10 RNA polymerase in one or more tissues or cell types.

The term "transgene capable of expressing" as used herein means a nucleic acid molecule comprising a nucleotide sequence encoding the polymerase operably linked to a promoter. The promoter may be any 15 promoter which functions in *C. elegans* and may be general (i.e. active in substantially all tissues and cell types), tissue-specific, cell type-specific, constitutive, inducible etc. Most preferably, the promoter will exhibit tissue or cell type-specificity.

20 With the use of a tissue or cell type-specific promoter of the appropriate specificity it is possible to control the site of RNA polymerase expression within *C. elegans* and hence control the site of expression of the recombinant polypeptide.

25 Methods for the construction of transgenic *C. elegans* worms are known in the art and are particularly described by Craig Mello and Andrew Fire, Methods in Cell Biology, Vol 48, Ed. H.F. Epstein and D.C. Shakes, Academic Press, pages 452-480.

30 In a further aspect the invention provides a kit for use in recombinant expression of a polypeptide in *C. elegans*, the kit comprising an outron cloning construct, as described above, and optionally a supply 35 of *C. elegans* nematode worms expressing an RNA polymerase specific for the bacteriophage promoter

- 12 -

present in the said outron cloning construct in one or more tissues or cell types.

The kit might further contain control inserts and control constructs, e.g. a reporter gene inserts and constructs which could be used to check efficiency of cloning steps and transfection steps, respectively. It might also contain constructs which may be used as selectable markers in the transfection procedure, e.g. a rol 6 plasmid (see below).

The invention further provides methods for the construction of transgenic *C. elegans* expressing a recombinant polypeptide in one or more tissues or cell types. One such method comprises introducing an outron expression construct, as described above, said construct being an expression vector suitable for use in *C. elegans* comprising an open reading frame encoding the desired recombinant polypeptide, into a *C. elegans* strain which expresses an RNA polymerase specific for the bacteriophage promoter present in said DNA construct in one or more tissues or cell types, and isolating transgenic *C. elegans* lines which stably express the said polypeptide. The *C. elegans* strain expressing the polymerase is preferably a transgenic strain carrying a transgene capable of expressing the RNA polymerase in one or more tissues or cell types, as described above. As aforesaid, transgenic *C. elegans* lines can readily be constructed using standard techniques well known in the art.

In an alternative approach, the method may comprise introducing into a background *C. elegans* strain (i) an outron expression construct, as described above, said construct being an expression vector suitable for use in *C. elegans* comprising an open frame encoding the desired recombinant polypeptide, and (ii) a DNA construct suitable for expression of an RNA polymerase specific for the

- 13 -

bacteriophage promoter present in the outron expression construct in one or more tissues or cell types of *C. elegans*, and isolating transgenic *C. elegans* lines which stably express the said 5 polypeptide. The second DNA construct may, advantageously, be an expression vector comprising a nucleotide sequence encoding the polymerase operably linked to a promoter having the appropriate tissue or cell type specificity.

10 In carrying out the methods of the invention one may employ standard techniques well known in the art for construction and selection of transgenic *C. elegans* lines. Such techniques are described, for example, in techniques described in *Methods in Cell 15 Biology*, vol 84; *Caenorhabditis elegans: modern biological analysis of an organism*, ed. Epstein and Shakes, academic press, 1995. Foreign DNA (e.g. plasmid DNA) may be introduced into *C. elegans* using microinjection or ballistic transformation, as 20 described in the applicant's co-pending International patent application No. WO 99/49066. In order to facilitate the selection of transgenic strains a marker plasmid may be co-introduced with the transgenes. A typical example is the plasmid pRF4 25 (Mello, C. C. et al. EMBO J. 10, 3959-3970 (1991)) which carries the rol-6 gene. *C. elegans* expressing rol-6 can be identified by screening for the roller phenotype. Any other *C. elegans* dominant selectable phenotypic marker, of which there are many known in 30 the art, may be used to facilitate selection of transgenic lines. A useful example is green fluorescent protein (or any of the equivalent autonomous fluorescent proteins known in the art).

35 In a still further aspect the invention provides transgenic *C. elegans* worms which contain an outron expression construct, as described above, said

- 14 -

construct being an expression vector suitable for use
in *C. elegans*, and which further express an RNA
polymerase specific for the bacteriophage promoter
present in the outtron expression construct in one or
5 more tissues or cell types.

The present invention will be further understood
with reference to the following non-limiting Examples,
together with the accompanying drawings in which:

10 Figure 1 illustrates the construction of a T7-outtron-
GFP vector. (A) sequence of the synthetic outtron
produced by annealing oligonucleotides o-GN59 and o-
GN60. (B) summary of the strategy used to construct
15 vector pDW3124.

Figure 2 shows plasmid maps for pDW3123 (outtron
cloning vector) and pDW3124 (outtron expression vector
for GFP expression).

20 Figure 3 is a plasmid map of pGN148 which contains a
T7 RNA polymerase coding sequence under the regulation
of the *C. elegans* SERCA promoter.

25 Figure 4 illustrates the nucleotide sequence of
pGN148.

Figure 5 illustrates the nucleotide sequence of pDW
3123 annotated to show the positions of the T7
30 promoter, outtron, synthetic intron A, multi-cloning
site and unc-54 3' UTR sequences and also the
ampicillin resistance gene.

Figure 6 illustrates the nucleotide sequence of pDW
35 3124 annotated to show the positions of the T7
promoter, outtron, synthetic intron A, GFP with introns
and unc-54 3' UTR sequences and also the ampicillin

- 15 -

resistance gene.

Example 1 -Construction of a T7-outron-GFP containing vector (pDW3124)

5 A SL1 trans-splice acceptor site (outron) was cloned into a vector downstream of the T7 promoter and upstream of the GFP to be expressed.

10 A synthetic outtron consisting of two partially overlapping oligonucleotides (o-GN59 and o-GN60, see Figure 1) was inserted into a XbaI/XmaI digested T7 promoter GFP construct. Briefly, 25 μ l o-GN59 and 25 μ l o-GN60 (100 μ M) were denatured for 5 minutes at 94°C, annealed for 30 minutes at 68°C then cooled to 4°C.

15 1 μ l of XmaI/XbaI digested pDW3120 and 10 μ l of the annealed oligos were then ligated using T4 ligase overnight at 16°C, transformed into competent *E. coli* and analysed by restriction digestion and DNA sequencing, all according to standard molecular biology procedures. The resulting vector was

20 designated pDW3124 (Figures 1 and 2).

The outtron contains an AU rich sequence followed by a splice-acceptor site as described by Conrad et al, NAR 21:913-919 (1993) (see Figure 1).

25 Example 2-Construction of a T7-Outron MCS vector

A general purpose vector was constructed to facilitate expression of other DNA sequences in *C. elegans* under the control of the T7 promoter. This was done by digesting vector pDW3124 with HindII (position 179) and PvuII (position 1029) (partial digest) and re-ligating the blunt ends, resulting in vector pDW3123 (Figure 2).

- 16 -

Example 3-The expression of heterologous genes in *C. elegans* regulated by the T7 promoter requires trans-splicing.

Wild-type *C. elegans* nematodes were co-injected
5 with various combinations of the following test
plasmids:

- 1) GFP reporter plasmid
GFP: pDW2020
10 outron-GFP: pDW2024
T7 promoter-GFP: pDW3120
T7 promoter-outron-GFP: pDW3124
- 2) T7 polymerase expression plasmid SERCA T7
15 polymerase: pGN148 together with pRF-4 (rol-6) as
marker.

For every co-injection experiment, a total concentration of 200 ng DNA/ μ l was used (plasmid 20 concentration was 50 ng/ μ l and carrier DNA was added up to 200ng/ μ l). For every co-injection ±15 adult worms were injected.

F1 offspring showing the marker rol-6 phenotype 25 were isolated and then selected for further study. The next generation (F2) of the roller lines were screened for GFP expression in the pharynx, vulva, tail and body wall muscles. These are the tissues in which the bacteriophage T7 RNA polymerase is known to 30 be expressed when under the control of the *C. elegans* SERCA promoter (as in the construct pGN148)

The results are shown in Table 1 below, which indicates the number of lines expressing GFP vs total number of lines isolated.

35

- 17 -

	1	2	3
A	Construct	no T7-polymerase construct	with T7-polymerase construct (50ng) pGNI48
B	GFP (50ng) pDW2020	0/8	2/6*
C	outron::GFP (50ng) pDW2024	0/11	3/8*
D	T7-promoter::GFP (50ng) pDW3120	0/3	0/5
E	T7-promoter::outron::GFP (50ng) pDW3124	0/7	13/13

* GFP-expression most probably result of recombination
 10 in the extrachromosomal array

No GFP expression was observed in the experiments
 15 where the T7 RNA polymerase was absent (cells B2, C2,
 D2, E2).

In the experiments where the T7 RNA polymerase
 expressing vector was co-injected with GFP vectors
 without a T7 promoter, as in the cells B3 and C3, GFP
 expression was sometimes observed. This is probably
 20 due to recombination events in the extrachromosomal
 arrays, resulting in transcription of GFP directly
 from the SERCA promoter.

In the experiments where the T7 promoter-GFP construct
 25 and the SERCA T7 RNA polymerase were co-injected, no
 GFP expression could be observed (cell D3). In
 contrast, all of the lines isolated from the
 experiments where the GFP transcript contained an
 outron at its 5' site (n=13) expressed GFP (cell E3).
 30 The outron is a favourable target for SL1
 trans-splicing. Since SL1 RNA molecules contain a 5'

- 18 -

trimethylguanosine CAP structure which is transferred to the mature mRNA this results in improved translation of the RNA and hence better expression of GFP. Without the outron the T7 RNA polymerase 5 transcripts do not carry a CAP structure at their 5' end, leading to inefficient translation. The results of this experiment illustrate the importance of trans-splicing for efficient expression of heterologous and homologous genes transcribed by 10 prokaryotic polymerases in *C. elegans*.

SEQUENCE LISTING

SEQ ID NO: 1 Oligonucleotide o-GN59
15 SEQ ID NO: 2 Oligonucleotide 0-GN60
SEQ ID NO: 3 Plasmid pDW3123
SEQ ID NO: 4 Plasmid pDW3124
SEQ ID NO: 5 Plasmid pGN148

- 19 -

Claims:

1. A DNA construct comprising a bacteriophage promoter operably linked to an outtron sequence.

5

2. A DNA construct as claimed in claim 1 which further comprises at least one restriction enzyme recognition site positioned downstream of and proximal to the outtron sequence.

10

3. A DNA construct as claimed in claim 2 which comprises a multi-cloning site positioned downstream of and proximal to the outtron sequence.

15

4. A DNA construct as claimed in claim 2 or claim 3 which further comprises a DNA fragment inserted at the said restriction site or at a restriction site within the said multi-cloning site.

20

5. A DNA construct as claimed in any one of claims 1 to 4 which is a replicable cloning vector.

25

6. A DNA construct as claimed in any one of claims 1 to 5 wherein the outtron sequence comprises a 3' splice acceptor site having the sequence TTTCAAG preceded by an AT-rich region.

30

7. A DNA construct as claimed in claim 6 wherein the outtron sequence comprises the nucleotide sequence illustrated in Figure 1A.

35

8. A DNA construct as claimed in any one of claims 1 to 7 wherein the bacteriophage promoter is the T7, T3 or SP6 promoter.

9. A DNA construct for use in bacteriophage promoter-driven expression of a polypeptide in a

- 20 -

eukaryotic host cell or organism, which construct comprises a bacteriophage promoter operably linked to a DNA sequence such that it is capable of initiating transcription of said DNA sequence upon binding of the 5 appropriate RNA polymerase to the promoter, wherein the said DNA sequence comprises an outtron sequence and at least one open reading frame positioned downstream of the outtron sequence.

10 10. A DNA construct as claimed in claim 9 which is an expression vector.

15 11. A DNA construct as claimed in claim 9 or claim 10 wherein the outtron sequence comprises a 3' splice acceptor site having the sequence TTTCAG preceded by an AT-rich region.

20 12. A DNA construct as claimed in claim 11 wherein the outtron sequence comprises the nucleotide sequence illustrated in Figure 1A.

25 13. A DNA construct as claimed in any one of claims 9 to 12 wherein the bacteriophage promoter is the T7, T3 or SP6 promoter.

14. A kit for use in recombinant expression of a polypeptide in *C. elegans*, the kit comprising a DNA construct as claimed in any one of claims 1 to 3, and optionally *C. elegans* worms expressing an RNA polymerase specific for the bacteriophage promoter present in said DNA construct in one or more tissues or cell types.

30 15. A method for expressing a recombinant polypeptide in *C. elegans* which method comprises: introducing a DNA construct as claimed in any one

- 21 -

of claims 9 to 13, said construct being an expression vector suitable for use in *C. elegans*, into a *C. elegans* strain which expresses an RNA polymerase specific for the bacteriophage promoter present in 5 said DNA construct in one or more tissues or cell types.

16. A method of generating transgenic *C. elegans* expressing a recombinant polypeptide, which method 10 comprises:

introducing a DNA construct as claimed in any one of claims 9 to 13 comprising an open reading frame encoding the recombinant polypeptide, said construct being an expression vector suitable for use in *C. elegans*, into a *C. elegans* strain which expresses an RNA polymerase specific for the bacteriophage promoter present in said DNA construct in one or more tissues 15 or cell types, and

isолating transgenic *C. elegans* lines which 20 stably express the said polypeptide.

17. A method of generating transgenic *C. elegans* expressing a recombinant polypeptide, which method comprises:

25 introducing into *C. elegans* (i) a first DNA construct as claimed in any one of claims 9 to 13 comprising an open reading frame encoding the recombinant polypeptide, said construct being an expression vector suitable for use in *C. elegans*, and 30 (ii) a second DNA construct suitable for expression of an RNA polymerase specific for the bacteriophage promoter present in the first DNA construct in one or more tissues or cell types of *C. elegans*, and

35 isolating transgenic *C. elegans* lines which stably express the said polypeptide

- 22 -

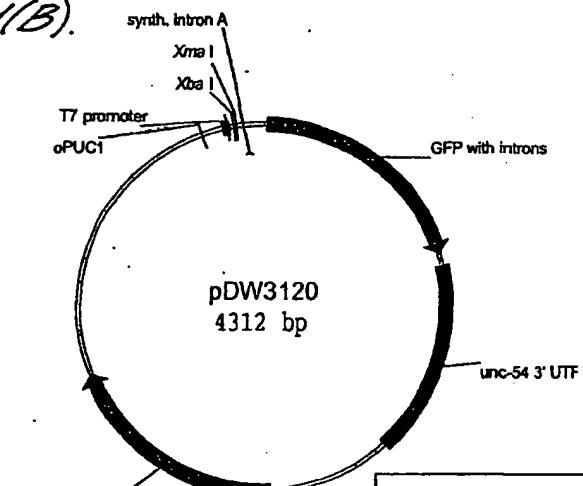
18. Transgenic *C. elegans* which contain a DNA construct as claimed in any one of claims 9 to 13, said construct being an expression vector suitable for use in *C. elegans*, and which further express an RNA polymerase specific for the bacteriophage promoter present in said DNA construct in one or more tissues or cell types.

5

FIG. 1(A).

XbaI overhang	SspI	3' splice acceptor
<u>CTAGATTACA</u> ACTAATTATA <u>CTTATTTGAA</u> TATT <u>CAAA</u> TTTC <u>CAGAC</u>		o-GN59
TAAT <u>GTTGATTA</u> ATAT <u>GAA</u> TA <u>AACT</u> TATA <u>AGTT</u> AAA <u>AGT</u> C <u>TGGGCC</u>		o-GN60
		XbaI overhang

FIG. 1(B).



XbaI/XbaI digested pDW3120
 25 µl o-GN59 + 25 µl o-GN60 (100 µM)
 Denature oligos o-GN59 & o-GN60 5 min. at 94°C
 Renaturate 30 min. at 68°C, cool to 4°C

Ligate 1 μ l vector + 10 μ l oligos with T4 ligase
Overnight at 16°C
Transform in E. coli
Analyse by Restriction Digest and sequencing

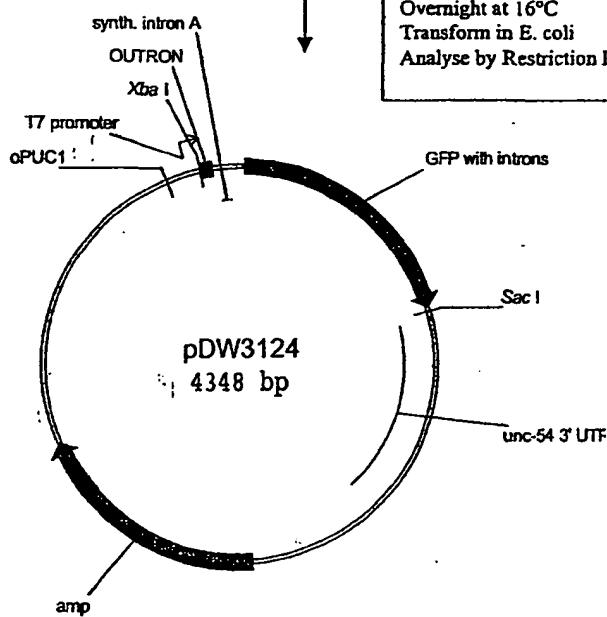


FIG. 2.

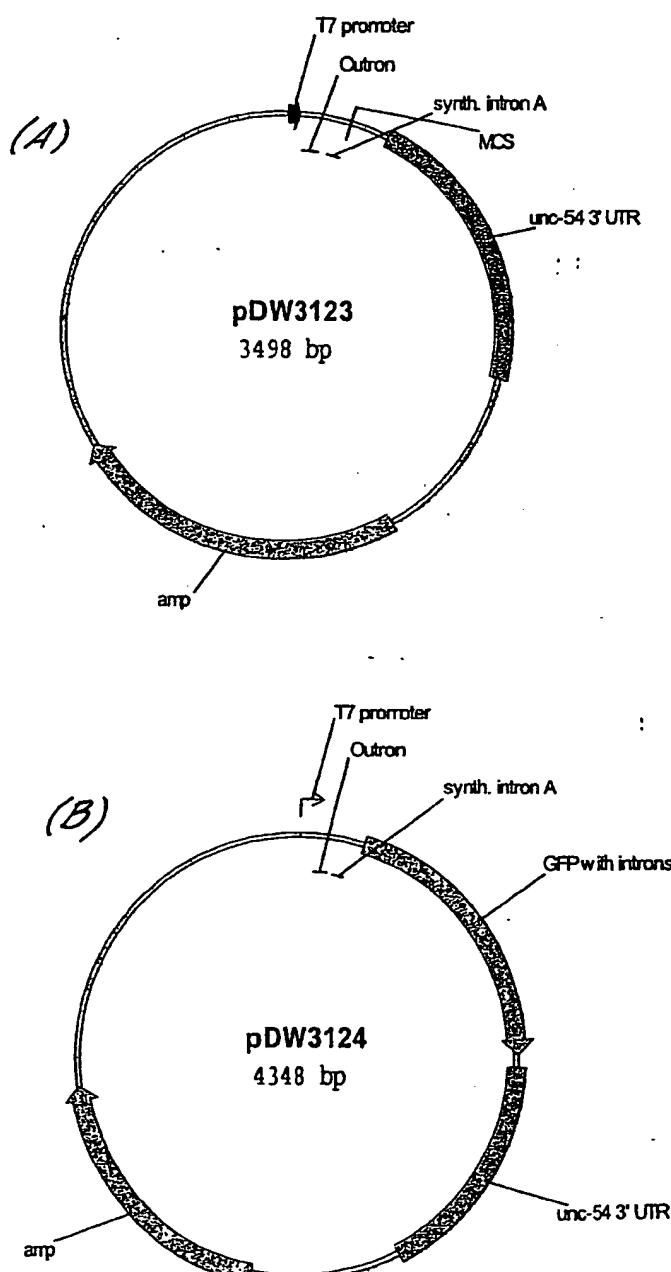


FIG. 3.

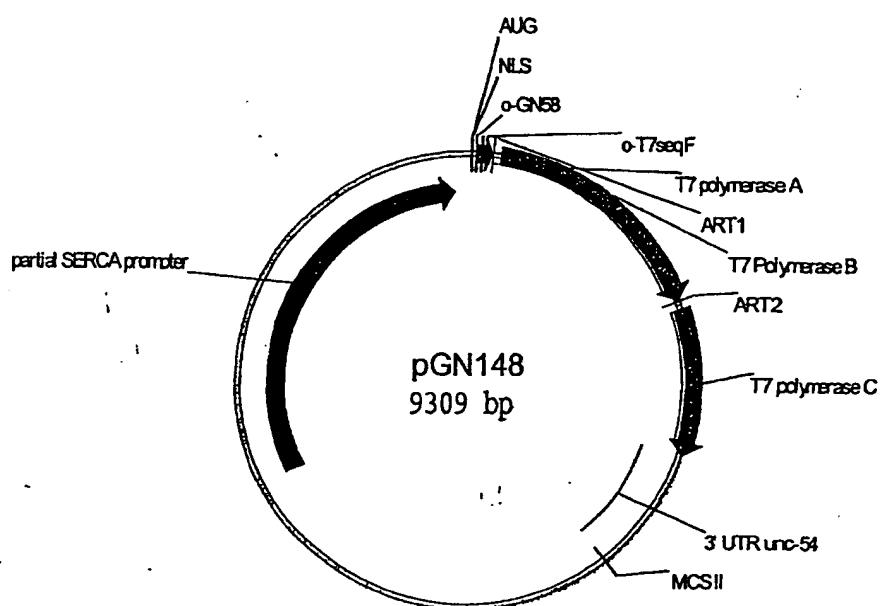


FIG. 4.

Nucleotide sequence of pGN148

FIG. 4 (CONTINUED 1.)

FIG. 4 (CONTINUED 2.)

FIG. 5.

T7 promoter

Outron

1 AGCTTGGCGC CTAATACGAC TCACTATAGG GCTGCAGGTC GACTCTAGAT TACAACATAAT TATACTTATT
TCGAACCGCG GATTATGCTG AGTGATATCC CGACGCCAG CTGAGATCTA ATGTTGATTA ATATGAATAA

Outron synth. intron A

71 TGAATATTCA AATTTTCAGA CCCGGGATTG GCCAAAGGAC CCAAAGGTAT GTTTCGAATG ATACTAACAT
ACTTATAAGT TTAAAAGTCT GGGCCCTAAC CGGTTCCCTG GGTTTCCATA CAAAGCTTAC TATGATTGTA

synth. intron A MCS

141 AACATAGAAC ATTTTCAGGA GGACCCCTGG CTAGCTCCT GCTGGGATTA CACATGGCAT GGATGAACTA
TTGTATCTG TAAAGTCTT CCTGGGAACC GATCGCAGGA CGACCTAAT GTGTACCGTA CCTACTTGT

unc-54 3' UTR

211 TACAATAGG GCGGGCCGAG CTCCGCATCG GCGCGCTGTCA TCAGATCGCC ATCTCGCGCC CGTGCCTCTG
ATGTTTATCC CGGCCGGCTC GAGGCGTAGC CGGCACAGT AGTCTAGCGG TAGAGCGCAG GCACGGAGAC

unc-54 3' UTR

281 ACTTCTAACTC CCAATTACTC TTCAACATCC CTACATGCTC TTCTCCCTG TGCTCCCACC CCCTATTTTT
TGAAGATTCA GGTTAATGAG AAGTTGTAAG GATGTACGAG AAAGAGGGAC ACAGGGGTGG GGGATAAAAAA

unc-54 3' UTR

351 GTTATTATCA AAAAAACTTC TTCTTAATTT CTTTGTTTTT TAGCTTCTTT TAAGTCACCT CTAACAATGA
CAATAATAGT TTTTTGAG AAGAATTTAA GAAACAAAAA ATCGAAGAAA ATTCACTGGG GATTGTTACT

unc-54 3' UTR

421 AATTGTGTAG ATTCAAAAT AGAATTAATT CGTAATAAAA AGTCGAAAAA AATTGTGTCTC CCTCCCCCCA
TTAACACATC TAAGTTTTA TCTTAATTAA GCATTATTT TCAGTTTTT TTAAACACGAG GGAGGGGGGT

unc-54 3' UTR

491 TTAAATAATAA TTCTATCCA AAATCTACAC AATGTTCTGT GTACACTTCT TATGTTTTT TTACTTCTGA
AATTATTATT AAGATAGGGT TTTAGATGTG TTACAAGACA CATGTGAAGA ATACAAAAAA AATGAAGACT

unc-54 3' UTR

561 TAAATTTTTT TTGAAACATC ATAGAAAAAA CCGCACACAA AATACCTTAT CATATGTTAC GTTCAGTTT
ATTTAAAAAA AACTTTGTAG TATCTTTTT GGCGTGTGTT TTATGAAATA GTATAACATG CAAAGTCAAA

unc-54 3' UTR

631 ATGACCGCAA TTTTATTTT TTGCGACGTC TGGGCCTCTC ATGACGTCAA ATCATGCTCA TCGTAAAAAA
TACTGGCGTT AAAAATAAG AACCCTGCAG ACCCGGAGAG TACTGCAGTT TAGTACGGAGT AGCACTTTTT

unc-54 3' UTR

701 GTTTGGAGT ATTTTTGGAA TTTTCATC AAGTGAAGT TTATGAAATT AATTTCTG CTTTTGCTTT
CAAACCTCA TAAAAACCTT AAAAAGTTAG TTCACCTTC AATACTTAA TAAAGGAC GAAACCGAAA

unc-54 3' UTR

771 TTGGGGGTTT CCCCTATTGT TTGTCAGAG TTGAGGAGC GGCGTTTTTC TTGCTAAAAT CACAAGTATT
AACCCCCAAA GGGGATAACA AACAGTTCTC AAAGCTCCTG CGCAGAAAG AACGATTTA GTGTTCTAA

unc-54 3' UTR

841 GATGAGCACG ATGCAAGAAA GATCGGAAGA AGGTTGGGT TTGAGGCTCA GTGGAAGGTG AGTAGAAGTT
CTACTCGTGC TACGTTCTT CTAGCCTCT TCCAAACCCA AACTCCGAGT CACCTCCAC TCATCTTCAR

unc-54 3' UTR

911 GATAATTGA AAGTGGAGTA GTGCTATGG GGTTTGCC TAAATGACA GAATACATTC CCAATATACC
CTATTAACCTT TTCACCTCAT CACAGATACC CCAAAACGG AATTACTGT CTTATGTAAG GGTTATATGG

unc-54 3' UTR

981 AAACATAACT GTTTCTACT AGTCGGCCGT ACGGGCCCTT TCGTCTCCGCG CGTTTGGTG ATGACGGTGAA
TTTGTATTGA CAAAGGATGA TCAGCCGGCA TGCCCGGAA AGCAGAGCGC GCAAAGCCAC TACTGCCACT

FIG. 5 (CONTINUED 1.)

1051 AACACCTCTGA CACATGCAGC TCCCCGAGAC GGTCACAGCT TGTCTGTAAG CGGATGCCGG GAGCAGACAA
TTTGGAGACT GTGTACGTCG AGGGCCTCTG CCAGTGTCA ACAGACATTC GCCTACGGCC CTCGTCTGTT

1121 GCGCGTCAGG GCGCGTCAGC GGGTGTGCGC GGCGTGCAGG GCTGGCTAA CTATGCGCA TCAGAGCAGA
CGGGCAGTCC CGCGCAGTCG CCCACAACCG CCCACAGCCC CGACCGAATT GATAACGCCGT AGTCTCGCT

1191 TTGTACTGAG AGTGCACCAT ATGCCGTGTA AAATACCGCA CAGATGCGTA AGGAGAAAAT ACCGCATCAG
AACATGACTC TCACGTGGTA TACGCCACAC TTTATGGCGT GTCTACGCAT TCCTCTTTTA TGGCGTAGTC

1261 GCGGCCCTAA GGGCCTCGTG ATACGCCAT TTTTATAGGT TAATGTCTG ATAATAATGG TTTCTTAGAC
CGCCGGAATT CCCGGAGCAC TATGCGGATA AAAATATCCA ATTACACTAC TATTATTACC AAAGAACCTG

1331 GTCAGGTGGC ACTTTTCGGG GAAATGTGCG CGGAACCCCT ATTGTTTAT TTTTCTAAAT ACATTCAAAT
CAGTCCACCG TGAAAGCCC CTTCACCGC GCCTGGGGA TAAACAAATA AAAAGATTTA TGTAAAGTTA

amp

1401 ATGTATCCGC TCATGAGACA ATAACCCCTGA TAAATGCTTC ATAATAATGG AAAAGGAAG AGTATGAGTA
TACATAGGGC AGTACTCTGT TATTGGACT ATTTACGAAG TTATTATAAC TTTTCCTTC TCATACTCAT

amp

1471 TTCAACATTTC CCGTGTGCC CTTATCCCT TTTTGCAGC ATTTGCCTT CCTGTTTTG CTCACCCAGA
AAGTTGAAA GGCACAGCGG GAATAAGGG AAAAAACGCCG TAAACCGGAA GGACAAAAC GAGTGGCT

amp

1541 AACGCTGGTG AAAGTAAAAG ATGCTGAAGA TCAGTTGGGT GCACGAGTGG GTTACATCGA ACTGGATCTC
TTGCGACAC TTTCACTTACGACTTCT AGTCAACCCA CGTGCCTCACC CAATGTAGCT TGACCTAGAG

amp

1611 AACAGCGGTAA AGATCCTTGA GAGTTTCGC CCCGAAGAAC GTTTCCAAT GATGAGCACT TTTAAAGTTC
TTGTCGCCAT TCTAGGAACT CTCARAGCG GGGCTCTTG CAAAGGTTA CTACTCGTGA AAATTCAAG

amp

1681 TGCTATGTGG CGGGTATTAA TCCCCTATTG ACGCCGGCA AGAGCAACTC GGTCGCCGCA TACACTATTC
ACGATACACC GCGCCATAAT AGGGATAAC TGCGGCCGT TCTCGTTGAG CCAGCGGGGT ATGTGATAAG

amp

1751 TCAGAATGAC TTGGTTGAGT ACTCACCAAGT CACAGAAAG CACCTTACGG ATGGCATGAC AGTAAGAGAA
AGTCTTACTG AACCAACTCA TGAGTGGTCA GTGTCTTTTC GTAGAATGCC TACCGTACTG TCATTCTCTT

amp

1821 TTATGCAGTG CTGCCATAAC CATGAGTGTAA AACACTGCCG CCAACTTACT TCTGACAACG ATCGGAGGAC
AATAACGTACAC GACGGTATTG GTACTCACTA TTGTGACGCC GGTTGAATGA AGACTGTGTC TAGCCTCTG

amp

1891 CGAAGGAGCT AACCGCTTT TTGCAACAAACG TGGGGGATCA TGTAACCTGC CTTGATCGTT GGGAAACCGGA
GCTTCCTCGA TTGGCGAAAA AACGTGTTGT ACCCCCTAGT ACATTGAGGG GAACTAGCAA CCCTTGGCCT

amp

1961 GCTGAATGAA GCCATACCAA ACGACGAGCG TGACACCACG ATGCCGTAG CAATGGCAAC AACGTTGCCG
CGACTTACTT CGGTATGGTT TGCTGCTCGC ACTGTGGTGC. TACGGACATC GTTACCGTTG TTGCAACGCC

FIG. 5 (CONTINUED 2.)

amp

2031 AACTATTAAC TGGCGAACT ACTTACTCTA GCTTCCCGC ACAAATTAAAT AGACTGGATG GAGGCCGATA
TTTGATAATT GACCGCTTGA TGAATGAGAT CGAAGGGCCG TTGTTAATTA TCTGACCTAC CTCCGCCTAT

amp

2101 AAGTTGCAGG ACCACTCTG CGCTCGGCC CTCGGCTGG CTGGTTTATT GCTGATAAAT CTGGAGCCGG
TTCAACGTCC TGGTGAAGAC GCGAGCCGGG AAGGCGACC GACCAAATAA CGACTATTTA GACCTCGGCC

amp

2171 TGAGCGTGGG TCTCGCGTA TCATTGCAGC ACTGGGGCCA GATGGTAAGC CCTCCCGTAT CGTAGTTATC
ACTCGCACCC AGAGGCCAT AGAACGTGG TGACCCCGGT CTACCATTCG GGAGGCCATA GCATCAATAG

amp

2241 TACACGACGG GGAGTCAGGC AACTATGGAT GAAAGAAATA GACAGATCGC TGAGATAGGT GCCTCACTGA
ATGTGCTGCC CCTCAGTCCG TTGATACCTA CTGCTTTAT CTGTCTAGCG ACTCTATCCA CGGAGTGACT

amp

2311 TTAAGCATTG GTAAGTGTCA GACCAAGTTT ACTCATATAT ACCTTAGATT GATTTAAAAC TTCAATTAAA
AATTGTAAC CATGACAGT CTGGTTCAAA TGAGTATATA TGAAATCTAA CTAATTGTTG AAGTAAAAT

2381 ATTTAAAAGG ATCTAGGTGA AGATCCTTT TGATAATCTC ATGACAAAAA TCCCTTAACG TGAGTTTTCG
TAAATTTCC TAGATCCACT TCTAGGAAA ACTATTAGAG TACTGGTTT AGGAAATTGC ACTCAAAGC

2451 TTCCACTGAG CGTCAGACCC CGTAGAAAAAG ATCAAAGGAT CTTCTTGAGA TCCCTTTTTT CTGCGCGTAA
AAGGTGACTC GCAGTCTGGG GCATCTTTG TAGTTCCCTA GAAGAACTCT AGGAAAAAAA GACGCGCATT

2521 TCTGCTGCTT GCAACACAAA AAACCACCGC TACAGCGGT GGTTGTTTG CCGGATCAAG AGCTACCAAC
AGACGACGAA CGTTGTTT TTTGGTGGCG ATGGTCGCCA CCAAACAAAC GGCTAGTTC TCGATGGTTG

2591 TCTTTTCCG AAGGTAACTG GCTTCAGCAG AGGCCAGATA CCAAATACTG TCCCTCTAGT GTAGCCGTAG
AGAAAAAGGC TTCCATTGAC CGAAGTCGTC TCGCGTCTAT GGTTATGAC AGGAAGATCA CTCGGCATC

2661 TTAGGCCACC ACTTCAAGAA CTCTGTAGCA CCCTCTACAT ACCTCGCTCT GCTAATCCG TTACCAAGTGG
AATCCGTGG TGAAGTCTT GAGACATCGT GGGGATGTA TGGAGCGAGA CGATTAGGAC AATGGTCACC

2731 CTGCTGCCAG TGGCGATAAG TCCTGTCTTA CCGGGTTGGA CTCAAGACGA TAGTTACCGG ATAAGGGCGCA
GACGACGGTC ACCGCTATTG AGCACAGAAAT GGGCCAACT GAGTCTGCT ATCAATGGCC TATTCCCGGT

2801 GCGGTCGGGG TGAACGGGGG GTTCGTGCAC ACAGCCCAGC TTGGAGCGAA CGACCTACAC CGAACTGAGA
CGCCAGCCCC ACTTCCCCC CAAGCACGTG TGTGGGTG AACCTCGCTT GCTGGATGTG CTTTGACTCT

2871 TACCTACAGC GTGAGCATTG AGAAAGCGCC ACCTTCCCG AAGGGAGAAA GGCGGACAGG TATCCGGTAA
ATGGATGTGG CACTCGTAAC TCTTTCGCGG TCGGAAGGGC TTCCCTCTT CCGCTGTC ATAGGCCATT

2941 GCGGCAGGGT CGGAACAGGA GAGGCCACGA GGGAGCTTC AGGGGGAAAC GCCTGGTATC TTTATAGTCC
CGCCGTCCA GCCTGTCTT CTCGCGTCTT CCCTCGAAGG TCCCCCTTG CGGACCATAG AAATATCAGG

3011 TGTCGGTTT CGCCACCTCT GACTTGAGCG TCGATTTTG TGATGCTGT CAGGGGGGG GAGCCTATGG
ACAGCCCCAA CGCGTGGAGA CTGAACCTCG ACAGAAAAAC ACTACGAGCA GTCCCCCCCCTCCTGGATAC

3081 AAAAACGCCA GCAACGCCGC CTTTTTACGG TTCTGGCCT TTTGCTGGCC TTTTGTCAAC ATGTTCTTTC
TTTTGCGGT CGTTGCCCG GAAAATGCC AAGGACCGGA AAACGACCGG AAAACGAGTG TACAAGAAAG

3151 CTGCGTTATC CCTGATTCT GTGGATAACC GTATTACCGC CTTTGAGTGA GCTGATACCG CTCGGCCGAG
GACGCAATAG GGGACTAAGA CACCTATTGG CATAATGGCG GAAACTCACT CGACTATGGC GAGCGCGTC

3221 CGGAACGACC GAGCGCAGCG AGTCACTGAG CGAGGAAGCG GAAGAGCGCC CAATACGCAA ACCGCCCTCTC
GGCTTGCTGG CTCGCGTCGC TCAGTCAC TCTCCTTCGC TTCTCGCGG GTTATGCGT TGGCGGAGAG

3291 CCCGCGCGTT GGCGATTCA TTATGCAGC TGGCACGACA GGTTTCCGA CTGGAAAGCG GGCAGTGAGC
GGGCGCGCAA CCGCTAACT AATTACGTGG ACCGTGCTGT CCAAAGGGCT GACCTTTCCGC CCGTCACTCG

3361 GCAACGCCAAT TAATGTGAGT TAGCTCACTC ATTAGGCACC CCAGGCTTTA CACTTTATGC TTCCGGCTCG
CGTTGCGTTA ATTACACTCA ATCGAGTGG TAATCCGTGG GGTCCGAAT GTGAAATACG AAGGCCGAGC

3431 TATGTTGTGT GGAATTGTGA GCGGATAACA ATTCACACA GGAAACAGCT ATGACCATGA TTACGCCA
ATACAACACA CCTAACACT CGCCTATTGT TAAAGTGTGT CTTTGTCGA TACTGGTACT AATGCGGT

FIG. 6.

	T7 promoter	Outtron
1	AGCTTGGCG CTAATACGAC TCACTATAGG GCTGCAGGTC GACTCTAGAT TACAACATAAT TATACTTATT TCCGACCCGG GATTATGCTG AGTGTATATCC CGACGTCACG CTGAGATCTA ATGTTGATTA ATAIGATAAA	Outtron
		synth. intron A
71	TGAATATTCA AATTTTCAGA CCCGGGATTG GCCAAAGGAC CCAAAGGTAT GTTTCGAATG ATACTAACAT ACTTATAAGT TAAAAGTCT GGGCCTAAC CGGTTCCCTG GGTTCCATA CAAAGCTAC TATGATTGTA	GFP with introns
		synth. intron A
141	AACATAGAC ATTTTCAGGA GGACCCCTGG CTAGCGTCGA CGGTACCATG GGGCGCGCCA TGAGTAAAGG TTGTATCTG TAAAAGTCCT CCTGGAAACC GATCGCAGCT GCATGGTAC CCCCGCGGT ACTCATTCTC	GFP with introns
		GFP with introns
211	AGAGGAACCT TTCACTGGAG TTGTCCCAAT TCTTGTGAA TTAGATGGTG ATGTTAATGG GCACAAATT TCTTCTTGAAGTGCACTC AACAGGGTAA AGAACAACTT AATCTACCCAC TACATTACCC CGTGTGTTAA	GFP with introns
		GFP with introns
281	TCTGCACTG GAGAGGGTGA AGGTGATGCA ACATACGGAA AACTTACCCCT TAAATTATT TGCACTACTG AGACAGTCAC CTCCTCCACT TCCACTACGT TGTATGCCCT TTGAATGGGA ATTAAATAAC ACGTGATGAC	GFP with introns
		GFP with introns
351	GAAACTACCC TGTCCCATGG GTAAGTTAA ACATATATAT ACTAACTAAC CCTGATTATT AAATTTCA CTTTGTGAGG ACAAGGTACG CATTCAAATT TGTATATATA TGATTGATTG GGACTAATAA ATTAAAGT	GFP with introns
		GFP with introns
421	GCCAAACACTT GTCACTACTT TCTGTTATGG TGTCAATGC TTCTCGAGAT ACCCAGATCA TATGAAACCG CGGTTGTGAA CAGTGATGAA AGACAATACC ACAAGTTACG AAGAGCTCTA TGGTCTAGT ATACTTTGCC	GFP with introns
		GFP with introns
491	CATGACTTT TCAAGAGTGC CATGCCGAA GTTATGTAC AGGAAAGAAC TATATTTTC AAAGATGACG GTACTGAAAA AGTCTCACG GTACGGGCTT CCATACATG TCCTTCTTG ATATAAAAG TTTCTACTGCC	GFP with introns
		GFP with introns
561	GGAACTACAA GACACGTAAG TTAAACACT TCGGTACTAA CTAAACCATAC ATATTTAAAT TTTCAGGTGCG CCTTGATGTT CTGTGATTC AAATTTGTC AGCCATGATT GATTGGTATG TATAAATTAA AAAGTCACCG	GFP with introns
		GFP with introns
631	TGAAGTCAGG TTGAGGGTG ATACCCCTGT TAATAGATTC GAGTTAAAG GTATTGATT TAAAGGAAGT ACTTCAGTTC AAACCTCCAC TATGGGAACA ATTATCTTAG CTCATTTTC CATAACTAA ATTCTCTCA	GFP with introns
		GFP with introns
701	GGAAACATTC TTGGACACAA ATTGGAAATAC AACTATACT CACACAATGT ATACATCATG GCAGACAAAC CCTTGTAAAG AACCTGTGTT TAACCTTAAG TTGATATTGA GTGTGTTACA TATGTTGATAC CGTCTGTTG	GFP with introns
		GFP with introns
771	AAAAGAATGG ATCAAAAGTT GTAAAGTTAA ACTTGGACTT ACTAACTAAC GGATTATATT AAATTTCA TTTTCTTACCTTACCA CATTCAAATT TGAACCTGAA TGATTGATTG CCTAAATATAA ATTAAAGT	GFP with introns
		GFP with introns
841	GAACCTCAGAA ATTAGACACA ACATTGAAAGA TGGAGCGTT CAACTACGAG ACCATTATCA ACAAAATACT CTTGTAAAGTT TAATCTGTGT TGTAACTCTC ACCTTCGCAA GTGTGATGTC TGCTAAAGT TGTTTATGA	GFP with introns
		GFP with introns
911	CCAATGGCG ATGGCCCTGT CCTTTTACCA GACAACCATC ACCTGTCCAC ACAATCTGCC CTTTCGAAG GGTTAACCGC TACCGGGACA GAAAATGGT CTGTTGGTAA TGGACAGGTG TGTTAGACGG GAAGCTTC	GFP with introns
		GFP with introns
981	ATCCCAACGA AAAGAGAGAC CACATGGTCC TTCTTGTGTT TGTAAACAGCT GCTGGGATTA CACATGGCAT TAGGGTTGCT TTCTCTCTG GTGTACAGG AAGAACTCAA ACATGTGCA CGACCCCTAA GTGTACCTA	GFP with introns

FIG. 6 (CONTINUED 1.)

GFP with introns

unc-54 3' UTR

1051 GGATGAACTA TACAAATAGG GCCGGCCGAG CTCCGCATCG GCGCCTGTCA TCAGATGCC ATCTCGCGCC
CCTACTTGAT ATGTTTATCC CGGGCGGCTC GAGGCGTAGC CGGGCACAGT AGTCTAGCGG TAGAGCGCGG

unc-54 3' UTR

1121 CGTGCCTCTG ACTTCTAAGT CCAATTACTC TTCAACATCC CTACATGCTC TTTCTCCCTG TGCTCCCACC
GCACGGGAGAC TGARGATTCA GGTTAATGAG AAGTTGTAGG GATGTACGAG AAAGAGGGAC ACCGAGGGTGG

unc-54 3' UTR

1191 CCCTATTTT GTTATTAATCA AAAAARCTTC TTCTTAATTY CTTTGTGTTT TAGCTTCTTT TAAGTCACCT
GGGATAAAA CAATAATAGT TTTTTGAAG AAGAATTAAA GAAACAAAAA ATCGAAGAAA ATTCAAGTGG

unc-54 3' UTR

1261 CTAACAATGA AATTTGTGAG ATTCAAAAAT AGAATTAATI CGTAATAAAA AGTCGAAAAA AATTTGTGCTC
GATTGTACT TTACACATC TAAGTTTTA TCTTAATTAA GCATTATTT TCAGCTTTT TTACACAGAG

unc-54 3' UTR

1331 CCTCCCCCCTA TTAAATAATAA TTCTATCCCA AAATCTACAC AAATGTTCTG GTACACTTCT TATGTTTTT
GGAGGGGGT AATTTATTAA AAGATAGGGT TTAGATGTG TTACAAGACA CATGTGAAGA ATACAAAAAA

unc-54 3' UTR

1401 TTACTCTGTA TAAATTTTT TTGAAACATC ATAGAAAAAA CGCACACAAA AATACCTTAT CATATGTTAC
AATGAAGACT ATTTAAAAAA AACTTTGTAG TATTTTTTT GCGCTGTGTT TTATGGAATA GTATACAATG

unc-54 3' UTR

1471 GTTTCACTTT ATGACCGCAA TTTTTATTC TTGCGACCGTC TGGGCTCTC ATGACGTCAA ATCATGCTCA
CAAAGTCAAA TACTGGCGTT AAAATAAAG AACGGTGCAG ACCGGAGAG TACTGCAGTT TACTACGACT

unc-54 3' UTR

1541 TCGTGAAGAA GTTTGGAGT ATTTTGGAA TTTTCAATC AAGTGAAGT TTATGAAATT AATTTCTG
AGCACTTTT CAAAACCTCA TAAAGCTTAA AAAAGTTAG TTCACTTCA AATACCTTAA TTAAAGGAC

unc-54 3' UTR

1611 CTTTGCTTT TTGGGGTTT CCCCTATTGT TTGCAAGAG TTGAGGAC GGCCTTTTC TTGCTAAAT
GAAACGAAA AACCCCCAAA GGGGATAACA AACAGTTCTC AAAGCTCTG CGCCTAAAG AACGATTTTA

unc-54 3' UTR

1681 CACAAGTATT GATGAGCAGG ATGCAAGAAA GATCGGAAGA AGGTTGGGT TTGAGGCTCA GTGGAAGGTG
GTGTTCAAA CTACTCGTGC TACGTTCTT CTACGCTTCT CCAAACCCA AACTCCGAGT CACCTTCCAC

unc-54 3' UTR

1751 AGTAGAAGTT GATAATTTGA AAGTGGAGTA GTGCTATGG GTTTTTGCTC TTAAATGACA GAATACATTC
TCATCTCAA CTATTAACCTTCACAGATACC CCAAAACGG AATTTACTGT CTTATGTAAG

unc-54 3' UTR

1821 CCAATATAACC AAACATAACT GTTCTCTACT AGTGGCCGT AGGGGCCCTT TCGCTCGCG CGTTTCCGGT
GGTTATATGG TTGTTATTGA CAAGGATGA TCAGCCGGCA TCGCCGGAA AGCAGAGCGC GCAAAGCCRC

1891 ATGACGGTGA AAACCTCTGA CACATGCAGC TCCCGGAGAC GTCACAGCT TGTCTGTAAG CGGATGCCGG
TACTGCCACT TTTGGAGCT GTGACGTGAG AGGGCCCTG CCAGTGTGCA ACAGACATTC GCCTACGGCC

1961 GAGCAGACAA GCCCGTCAGG GCGGCTCAGC GGGTGTGCGG GGGTGTGCGG GCTGGCTAA CTATGCCGCA
CTCGTCTGTT CGGGCAGTCC CGCGCAGTGC CCCACACCG CGACCGAATT GATAACGGCGT

2031 TCAGAGCAGA TTGTACTGAG AGTGCACCAT ATGCGGTGAG AAATACCGCA CAGATGCCA AGGAGAAAT
AGTCTCGTCT AACATGACTC TCACGTTGTA TACGCCACAC TTATGGCGT GTCTACGCCAT TCCTCTTTA

2101 ACCGCATAG CGGGCCTTAA GGGCCTCGTG ATACGCCAT TTATATAGGT TAATGTCATG ATAATAATGG
TGGCGTAGTC CGCCGGAAATT CCCGGAGCAC TATGCGGATA AAAATATCCA ATTACAGTAC TATTATTAC

FIG. 6 (CONTINUED 2.)

2171 TTTCTTAGAC GTCAGGTGGC ACTTTTGGG GAAATGTGGG CGGAACCCCT ATTGTTAT TTTCTAAAT
AAAGAACCTG CAGTCCACCG TGAAAAGCCC CTTTACACGC GCCTTGGGA TAACAAATA AAAAGATTTA

2241 ACATTCAAAAT ATGTATCCGC TCATGAGACA ATAACCCCTGA TAAATGCTTC AATAATATG AAAAGGAAG
TGTAAGTTA TACATAGGCG AGTACTCTGT TATTGGACT ATTTACGAAG TTATTATAAC TTTTCCCTTC

amp

2311 AGTATGAGTA TTCAACATTT CCGTGTGCC CTATTCCCT TTTTGCAGG ATTTGCCTT CCTGTTTTG
TCATACTCAT AAGTTGAAA GGACAGCGG GAATAAGGG AAACACGCCG TAAACCGGAA GGACAAAAC

amp

2381 CTCACCCAGA AACGCTGGTG AAAGTAAAAG ATGGCTGAAGA TCAGTTGGGT GCACGAGTGG GTTACATCGA
GAGTGGGTCT TTGCGACCAC TTTCATTTC TACGACTTCT AGTCAACCCA CGTGCTCACC CAATGTAGCT

amp

2451 ACTGGATCTC AACAGCGGTA AGATCCTTGA GAGTTTCGC CCCGAAGAAC GTTTCCAAT GATGAGGACT
TGACCTAGAG TTGTCGCCAT TCTAGGAAC TCTAAAGCG GGGCTTCTTG CAAAAGGTTA CTACTCGTGA

amp

2521 TTTAAAGTTG TGCTATGTTG CGCGGTATTA TCCCGTATTG ACGCCGGGCA AGAGCAACTC GGTCGCCGCA
AAATTCAGG ACGATAACCC GCGCCATAAT AGGGCATAAC TGCGGCCGT TCTCGTTGAG CCAGCGCGT

amp

2591 TACACTATTG TCAGAATGAC TTGGTTGAGT ACTCACCAGT CACAGAAAAG CATCTTACGG ATGGCATGAC
ATGTGATAAG AGTCTTACTG ACCAAACTCA TGAGTGGTCA GTGTCTTTG GTAGAATGCC TACCGTACTG

amp

2661 AGTAAGAGAA TTATGCAGTG CTGCCATAAC CATGAGTGAT AACACTGCGG CCAACTTACT TCTGACAACG
TCATTCTCTT AATACGTAC GACGGTATTG GTACTCACTA TTGTGACGCC GGTTGAATGA AGACTGTTGC

amp

2731 ATCGGAGGAC CGAAGGGAGCT AACCGCTTTT TTGCACAACA TGCGGATCA TGTAACTCGC CTTGATCGTT
TAGCCTCTG GCTTCTCGA TTGGCGAAA AACGTGTTGT ACCCCCTAGT ACATTGAGCG GAACTAGCAA

amp

2801 GGGAAACCGGA GCTGAATGAA GCCATACCAA ACACGAGGCG TGACACCACG ATGCCCTGAG CAATGGCAC
CCCTTGGCCT CGACTTACTT CGGTATGGTT TGCTGCTCGC ACTGTGGTGC TACGGACATC GTTACCGTTG

amp

2871 AACGTTGCGC AAACATTAA CTGGCGAACT ACTTACTCTA GCTTCCCGC ACAATTAAAG ACTGGATG
TTGCAACCGG TTGATAATT GACCGCTTGA TGAATGAGAT CGAAGGGCG TTGTTAATT TCTGACCTAC

amp

2941 GAGGCAGGATA AAGTTGCGAGG ACCACTCTG CGCTCGGCCCT TTCCGGCTGG CTGGTTTATT GCTGATAAT
CTCCGCCTAT TTCAACGTCC TGGTGAAGAC GCGAGCCGG AAGGCCGACC GACCARATAA CGACTATTAA

amp

3011 CTGGAGCCGG TGAGGGTGGG TCTCGCGGT ATCATTGAGC ACTGGGGCCA GATGGTAAGC CCTCCCGTAT
GACCTCGGCC ACTCGCACCC AGAGCGCCAT AGTAACGTUG TGACCCCGT CTACCATCG GGAGGGCATA

amp

3081 CGTAGTTATC TACACGACGG GGAGTCAGGC AACTATGGAT GAACGAAATA GACAGATCGC TGAGATAGGT
GCATCAATAG ATGTGCTGCC CCTCAGTCGG TTGATACTA CTTGCTTTAT CTGTCTAGCG ACTCTATCCA

FIG. 6 (CONTINUED 3.)

amp

3151 GCCTCACTGA TTAAGCATGG GTAACTGTCA GACCAAGTTT ACTCATATAT ACTTTAGATT GATTAAAAAC
CGGACTGACT AATTGTAAC CATTGACAGT CTGGTCAA TGAGTATATA TGAAATCTAA CTAATTTTG

3221 TTCACTTTA ATTTAAAAGG ATCTAGGTGA AGATCCTTT TGATAATCTC ATGACCAAAA TCCCTTAACG
AAGTAAAAAT TAAATTTCC TAGATCCACT TCTAGGAAA ACTATTAGAG TACTGGTTT AGGGAAATTGC

3291 TGAGTTTCG TTCACTGAG CGTCAGACCC CGTAGAAAAG ATCAAAGGAT CTTCTTGAGA TCCTTTTT
ACTCAAAGC AAGGTGACTC GCAGTCTGGG GCATCTTTC TAGTTCTTA GAAGAACTCT AGGAAAAAAA

3361 CTGCGCGTAA TCTGCTGCTT GCAAACAAA AAACCACCGC TACCAAGGGT GGTTTGTGTTG CCGGATCAAG
GACGCCATT AGACGACGAA CGTTGTTT TTTGGTGGCG ATGGTCGCCA CCAAACAAAC GGCCTAGTT

3431 AGCTACCAAC TCTTTTCCG AAGGTAACTG GCTTCAGCAG AGCCAGATA CCAAATACTG TCCTTCTAGT
TCGATGGTT AGAAAAAGGC TTCCATTGAC CGAAGTCGTC TCGCGTCTAT GGTTATGAC AGGAAGATCA

3501 GTAGCCGTAG TTAGGCCACC ACTTCAAGAA CTCTGTAGCA CCGCCTACAT ACCTCGCTCT GCTAATCCTG
CATGGCATIC AATCCGGTGG TGAAGTTCTT GAGACATCGT GGCGGATGTA TGGAGCGAGA CGATTAGGAC

3571 TTACCACTGG CTGCTGCCAG TGGCGATAAG TCGTGTCTTA CCGGGTTGGA CTCAGACGA TAGTTACCGG
AATGGTCACC GACGACGGTC ACCGCTATTIC AGCACAGAAAT GGCCCAACCT GAGTTCTGCT ATCAATGGCC

3641 ATAAGGGCA GCGGTCGGGC TGAACGGGGG GTTCGTGAC ACAGCCCAGC TTGGAGCGAA CGACCTACAC
TATTCCCGT CGCCAGCCCC ACTTGGCCCC CAAGCACGTG TGTGGGTCG AACCTCGCTT GCTGGATGTG

3711 CGAACTGAGA TACCTACAGC GTGAGCATTG AGAAAGCGCC ACGCTCCCG AAGGGAGAAA GGCGGACAGG
GCTTGAETCT ATGGATGTG CACTCGTAAC TCTTCGGGG TGCGAAGGGC TTCCCTCTT CCGCTGTCC

3781 TATCCGGTAA GCGGCAGGGT CGGAACAGGA GAGCGCACGA GGGAGCTTCC AGGGGAAAC GCGTGTATC
ATAGGCCATT CGCCGTCCCA GCCTTGTCTT CTCGCGTGT CCCTCGAAGG TCCCCCTTG CGGACCATAG

3851 TTTATAGTCC TGTCGGGTTT CGCCACCTCT GACTTGAGCG TCGATTTTTG TGATGCTCGT CAGGGGGCG
AAATATCAGG ACAGCCCCAA GCGGTGGAGA CTGAACCTCGC AGCTAAAC ACTACGAGCA GTCCCCCGC

3921 GAGCCATGG AAAAACGCCA GCAACCGGGC CTTTTACGG TTCCTGGCCT TTGCTGGCC TTTGCTCAC
CTCGGATACC TTTTGGCGT CGTTGGCGG GAAAATGCC AAGGACCGGA AACGACCGG AAAACGAGTG

3991 ATGTTCTTC CTGCGTTATC CCTGTATTCT GTGGATAACC GTATTACCGC CTTTGAGTGA GCTGATACCG
TACAAGAAAG GACGCAATAG GGGACTAAGA CACCTATTGG CATAATGGCG GAAACTCACT CGACTATGGC

4061 CTCGCCGAG CGAACGACCC GAGCGCAGCG AGTCAGTGG CGAGGAAGCG GAAGAGCGCC CAATACGCAA
GAGCGCGTC GGTTGCTGG CTCGCGTGC TCAGTCACTC GCTCCCTCGC CTTCTCGGG GTTATGCGTT

4131 ACCGCCCTCTC CCCGCGCGTT GGGCGATTCA TTAATGCAGC TGGCACGACA GGTTTCCCGA CTGGAAAGCG
TGGCGGAGAG GGGCGCGCAA CGGGCTAAGT AATTACGTGG ACCGTGCTGT CCAAAGGGCT GACCTTCGC

4201 GGCAGTGAGC GCAACGCAAT TAATGTGAGT TAGCTCACTC ATTAGGCACC CCAGGCTTAA CACTTATGC
CCGTCACTCG CGTTGCGTA ATTACACTCA ATCGAGTGG TAAATCCGGG GGTCCGAAT GTGAAATAGC

4271 TTCCGGCTCG TATGTTGTGT GGAATTGTGA GCGGATAACA ATTTCACACA GGAAACAGCT ATGACCATGA
AAGGCCGAGC ATACAACACA CCTTAACACT CGCCTATGT TAAAGTGTGT CCTTGTGCA TACTGGTACT

4341 TTACGCCA
AATGCGGT

SEQUENCE LISTING

<110> DEVGEM NV

<120> GENE EXPRESSION SYSTEM

<130> SCB/55177/001

<140>

<141>

<160> 5

<170> PatentIn Ver. 2.0

<210> 1

<211> 47

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:

oligonucleotide o-GN59

<400> 1

ctagattaca actaattata ctatggaa tattcaaatt ttcatagac

47

<210> 2

<211> 47

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:

oligonucleotide o-GN60

<400> 2

ccgggtctga aaatttgaat attcaaataa gtataattag ttgtat

47

<210> 3

<211> 3498

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: plasmid

pDW3123

<400> 3

agcttggcgc ctaatacgac tcactatagg gctgcagggtc gactctagat tacaactaat 60
tatacttatt tgaatattca aattttcaga cccgggattt gccaaaggac ccaaaggat 120
gtttcgaaat atactaacat aacatagaac attttcaggaa ggacccttgg ctatcgatct 180
gttgggatata cacatggcat ggatgaacta tacaatagg gcccggccag ctccgcacatcg 240
ggccgtgtca tcagatcgcc atctcgcc cgtgcctctg acttctaaatg ccaattactc 300
ttcaacatcc ctacatgctc ttctccctg tgctccccacc ccctatTTT gttattatca 360
aaaaaaatcc ttcttaatTTT ctttgttttt tagttcttt taagtccatct ctaacaatga 420
aattgtgttag attcaaaaat agaattaatt cgtaataaaaa agtcgaaaaa aatttggtctc 480
ccccccttcca ttaataataa ttctatccca aatctcacac aatgttctgt gtacacttct 540
tatgtttttt ttacttctga taaaatTTT ttgaacatc atagaaaaaa ccgcacacacaa 600
aatacccttat catatgttac gttcagttt atgaccgcaa ttTTTatTC ttccgcacgtc 660
tgggcctctc atgaegtcaa atcatgctca tcgtaaaaaa gttttggagt atttttggaa 720

<210> 4
<211> 4348

<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: plasmid
pDW3124

E4000> 4

```
4<60> 4  
agcttggc ctaatacgac tcactatagg gctcgaggc gactcttagat tacaactaat 60  
tatactttt tgaatattca aattttcaga cccgggattt gccaaaggac ccaaaggat 120  
gtttcgaatg atactaacat aacatagaac attttcagga ggacccttgg cttagctcga 180
```

cggttacccat gggcgcgcca ttagtaaagg agaagaactt ttcaactggag ttgttccaaat 240
tcttgttcaa ttagatggg atgttaatgg gcacaattt tctgtcagt gagaggggtga 300
aggtgatgca acatacggaa aacttaccct taaatttttactgactaactg gaaaactacc 360
tgttccatgg gtaagttaa acatatatat actaactaac cctgattttaaattttca 420
gccaacactt gtcaactact tctgttatgg tggttcattgc ttctcgagat acccgatca 480
tatgaaacgg catgacttt tcaagagtgc catgcccggaa gggttatgtc agggaaaaac 540
tatattttt aaagatgacg ggaactacaa gacacgtaaat tttaaacagt tcggactaa 600
ctaaccatc attttttat ttcagggtgc tgaactcaag tttgaaggtg atacccttg 660
taatagaatc gaggtaaaag gtattgattt taaaaggat gggaaacattt ttggacacaa 720
attggaaatc aactataact cacacaatgt atacatcatg gcagacaaac aaaagaatgg 780
aatcaaaatg gtaagttaa acttggactt actaactaac ggattatatt taaattttca 840
gaacttcaaa attagacaca acattgaaga tggaaagcgtt caactagcag accattatca 900
acaatataact ccaattggcg atggccctgt ctttttacca gacaaccattt acctgtccac 960
acaatctgcc ctggcggaa atcccaacga aaagagagac cacatggtcc ttcttgatgtt 1020
tgtaacagct gctgggatta cacatggcat ggttggacta tacaatagg gcccggccag 1080
ctccgcctcg gccgcgtca tcagatgcct atctcgccg cgtgcctgt actttaatgt 1140
ccaatatttc ttcaacatcc tcatactgtc ttctccctgt tgctccacc cccttatttt 1200
gttattttca aaaaacttc ttcttaattt tttgttttt tagtttttt taatgtcacct 1260
ctaacaatga aattgtgttag attcaaaaat agaattaattt cgtaataaaa agtgcaaaaaa 1320
aattgtgtcc cttcccccca ttaataataa ttctatccca aatctcacac aatgttctgt 1380
gtacacttct tatgtttttt ttacttctga taaattttttt ttgaacatc atagaaaaaa 1440
ccgcacacaa aataccat tcatatgttac gtttcagttt atgaccgaa ttttttatttc 1500
ttcgacacgtc tgggccttc atgacgttca atcatgtctca tctgtgaaaaa gttttggagt 1560
atttttggaa tttttcaatc aagtggaaat ttagttaattt aattttctgt cttttgtttt 1620
ttgggggttt cccctattgt ttgtcaagag ttgcaggag ggcgtttttc tgctaaaaat 1680
cacaaggatt gatgagcagc atgcaagaaaa gatcggaga ggttgggtt ttgaggctca 1740
gtggaaagggt agtagaaatgtt gataatttga aagtggagta ttgtctatgg gtttttgc 1800
ttaaatgaca gaatacatc ccaatatacc aaacataact gtttctact agtggccgt 1860
acggggccctt tcgtctcgcg cgtttccgtg atgacgggtt aaacctctga cacatgcgc 1920
tccccggagac ggtcacacgt tgctgttaag cggatgcccgg gaggacacaa gcccgtcagg 1980
gcccgtcagc ggggttggc ggggttgcggg gttggcttaa ctatgcggca tcagacgaga 2040
ttgtacttag agtgcacat atgcgggttca aataatccgc cagatgcgtt agggaaaaat 2100
accgcatcg gggcccttaa gggctctgtg atacgcctat tttttaatgtt taatgtcatg 2160
ataataatgg tttcttagac tcacatgttgc acttttccggg gaaatgttgc cggaaaccct 2220
atttttttat ttttctaaat acattcaat atgtatccgc tcatgagaca ataaccctga 2280
taaatgttcc aataatattt gaaaaggaaag agttagatgtt ttcaacatcc cctgtcgcc 2340
cttattccct ttttgcggc attttgcctt cctgtttttt ctacccaga aacgtgggt 2400
aaatggaaat atgctgttca tcaatgttgc gcaatgggtt gttacatcgacttgc 2460
aacacgggtt agatccttgc gaggatccgc cccggaaagac gtttcccaat gatgagact 2520
tttaaaggttc tgctatgttgc cgggttataa tcccgatggt acgcggggca agagaacact 2580
ggtcggccca tacatatttgc tcaaatgttca ttgtgttgcgtt acttcaccaatg cacaaggaaat 2640
catcttacgg atggcatgac agtaagagaa ttatgcgttgc ctgcataacatcatgttgc 2700
aacactgcgg ccaacttact tctgacaacgc atcggaggac cgaaggagct aaccgccttt 2760
ttgcacaaca tggggatca tggtaactcgc ttgtatcggtt gggaaaccggaa gctgaatgaa 2820
gccccatccaa acgacgagcg tgacaccacg atgcctgttca atatggcaac aacgttgcgc 2880
aaacttattaa ctggcgaaat acttactcta gtttccggc aacaatttactt agactgttgc 2940
gaggcggtt aagtgcggg accacttgc ctgcgtggcc ttccggctgg ctgggtttttt 3000
gctgatataat ctggcgccgg tgacgttggg tctegcggtt tcatgtcgacttgcggcc 3060
gatggtaagc cttccatgttca tggatgttca tacaacgcgg gggatgttgc acatgttgc 3120
gaacggaaaat gacatgttgc tgatgttgcgtt gcttgcgttca ttaaaggatgttca 3180
gaccaatgtt acttataat ttttagattt gatggaaaat ttcatgttca atttttttt 3240
atcttaggttca agatcctttt tgataatctc atgaccaaaa tcccttaacg ttaggtttcg 3300
ttccactgttgc cgtcagaccc cttttttttt atcaaaaggat ttcttgcgtt tccctttttt 3360
ctggcggttca tctgtgttca gcaacaaaaaa aaccacccgc taccagcggtt gttttttttt 3420
ccggatcaag agtcaacaaat tctttttccg aaggtaacttgc gtttcagatgc agcgcagata 3480
ccaaatatactg tccttctgttca tggccgtt gttttttttt tcccttaacg ttaggtttcg 3540
ccgcctatcat acctcgctt cttttttttt taccatgttgc tctgtcgccg tggcgatata 3600
tcgtgttcaat cccgggttgc cttcaagacgc tttttttttt ataaaggccca ggggttggcc 3660
tgaacggggg gtttgcgttca acagccacgc ttggagcgaa cttttttttt tttttttttt 3720
tacccatcgatc gtggatgttca agaaagcgcc acgttcccg aaggagaaaaa gggggacagg 3780
tatccggtaa cggccgggttca gggaaacaggat gggggatgttca tttttttttt 3840

```

ggctggatc ttatacgatcc tgcgggttt cgccacactt gacttgagcg tggattttg 3900
tggatcgatc caggggggcg gagcctatgg aaaaacgcca gcaacgcgc ctttttacgg 3960
tttcgtggctt ttgtcgcc ttttgcac atgttcttc ctgcgttac ccctgattct 4020
gtggataacc gtattaccgc ctggaaatgtga gctgataccg ctgcggcgcg ccgaaccgacc 4080
gagcgcgcgc agtcaatgtgg cgagaaagcg gaagagcgcc caatacgcaa accgcctctc 4140
ccccgcgcgtt ggccgatcca ttaatgcgcg tggcaca gggttcccgaa ctggaaagcg 4200
ggcgtgtgaa gcaaccaat taatgtgtatgcgt tagctcaatc attaggcacc ccaggcttta 4260
cactttatgc ttccggctcg tatgttgtgtt gaaattgtga gcccataaca atttcacaca 4320
ggaaacagct atgaccatga ttacgcca 4348

```

<210> 5
<211> 9309

<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: plasmid pGN148

<400> 5

atgactgtc caaagaagaa gcgttaaggta ccggtaatga acacgattaa catcgtaag 60
aaccgacttct ctgacatcga actggctgt atcccggtca acactctggc tgaccattac 120
ggtgagcgtt tagctcggtg agttttaaca tctagatact aactaaacgt taacattaa 180
attttcagcg aacagtgtgc ccttggcat gaggcttacg agatgggtga aacgcgttc 240
cgcaagatgt tttagctgtca acttaaagat ggtgagggtt cggataaegc tgccccaag 300
ctcttcatca ctatccctact ccctaagatg gtgcacgc tcaacgcgt tggtgaggaa 360
gtgaaagctg agcgcgcaaa gcccggaca gccttccagt tcctgcaaga aatcaagccg 420
gaagccgtag cgtagatcac cattaagacc actctggctt gcctaaccag tgctgacaaat 480
acaaccgttc aggctgttagc aagcgeaate ggtcgccca ttgaggacga ggctcgcttc 540
ggtcgtatcc gtgaccttga agctaaggac ttcaagaaaa acgttgagga acaactcaac 600
aagcgcgtag ggcacgtcta caagaagca ttatgcgaag ttgtcgaggg tgacatgctc 660
tctaagggtc tactcggtt cgaggctgg tcttcgtggc ataaggaaaga ctctattatc 720
gttaggtatc gtgcacgtca gatgtccat gaggtaaccg gagggtttag cttagaccgc 780
aaaatgtctg gctgttagg tcaagactt gagaactcg aactcgccatc tgaatcgct 840
gaggctatcg caaccggcgc aggtgcgtg gctggcatct tcggatgtt ccaaccgtc 900
gtagttccctc ctaagccgtg gactggcatt actgggtgtt gctattggc taacggctgt 960
cgccctctgg cgctggcgc tactcacagt aagaaagcac tgatgcgta cgaagacgtt 1020
tacatgcctg aggtgtacaa agcgattaaac attgcgcacaa acaccgcgt gaaaatcaac 1080
aagaaagtcc tagcggtcgc caacgtaaatc accaagtgg agcattgtcc ggtcgaggac 1140
atccctgcga tttagctgtg agaactcccc atggaaacccgg aagacatcga catgaatcc 1200
gaggctctca ccgcgtggaa acgtgcgtcc gttgtgtt accgcgaagg caaggctcgc 1260
aagtctcgcc gtatcgcct tgatgttcatg tggagcaag ccaataagg tgcataaccat 1320
aaggccatct ggttccctta caacatggac tggcgccgtc gtgttacgc tgcgtcaatg 1380
ttcaaccgc aagctaacga tatgacccaa ggactgtta cgctggcga aggttaaccga 1440
atcggttaagg aagggtacta ctggctgaaa atccacgggt ccaaactgtc ggggtgtcgat 1500
aagggtccgt tccctgagcg catcaagttc attgaggaaaa accacgagaa catcatggct 1560
tgcgctaagt ctccactgga gaacacttgg tgggctgagc aagattctcc gttctgttcc 1620
cttgcgttct gctttgagta cgctgggtt cagcaccacg gcctgagctt taactgtcc 1680
cttgcgttgg cggttgcacgg gtctgtctt ggcattccacg acttctccgc gatgtcccg 1740
gatgaggtag gtggcgcgc ggtttaatgt ttaactcta tcctactaac taacgaagct 1800
tatatttttttttcagaactt gtttcttagt gaaaccgttcc agacatcga cgggattgtt 1860
gctaagaaaat tcaacgagat tctacaagca gacgcatac atgggaccga taacgaagta 1920
gttaccgttca ccgttgcacggaa cactggtaaa atctctgaga aagtcaagct gggactaag 1980
gcactggctg gtcaatggct ggcttacgggt gttactcgca gtgtgactaa gcgttcgtc 2040
atgacgttgg cttacgggtc caaagatgtc ggcttccgtc aacaagtgtc ggaagatacc 2100
attcagccatc ctattgttcc cggcaagggtt ctgtatgtca ctcagccaa tcaggctgtc 2160
ggatcatgg ctaagctgtat ttggaaatct gtggcgttca cgggtgttgc tgccgttgg 2220
gcaatgaactt ggtttaatgtc tgctgtcaag ctgtgttgcgtt ctgggttcaaa agataagaag 2280
actggagata ttcttcgcaaa gctgttgcgtt gtcatttttttgc tcaactccggta tggttccct 2340
gtgtggcagg aatacaagaa gcttattcaaa acgcgttttga acctgtatgtt ctcgttgc 2400
ttccgcttac agccttccat taacaccaac aaagataacg agattgttgc acacaaacag 2460
gagtctggta tcgcttccaa ctttgtacac agccaaagacg gtagccacct tgcgttgc 2520

gtagtgtgg cacacgagaa gtacggaatc gaatctttt cactgattca cgactccccc 2580
 ggtaccatc cggctgacgc tgcgaaacctg ttcaaaggcag tgcgcaaaac tatgttgac 2640
 acatatgagt ctgtgtatgt actggctgt ttctacgacc agttcgctga ccagttgcac 2700
 gagtcataat tggacaaaat gccagcaactt cccgctaaag gtaacttgaa ctcctgtac 2760
 atcttagagt cggacttcgc gtcgcgtaa gaattccaaac tgagcggccg tcgctaccat 2820
 taccacttgc tctgggtgtca aaaataatag gggccgtgt catcagaga agttaaaact 2880
 gagttctact aactaacgag taatattaa atttcagca tctcgccccc tgccctgtca 2940
 cttctaagtc caattactct tcaacatccc tacatgtctt ttctccctgt gctcccaccc 3000
 cctattttt ttattatcaa aaaaactct tcttaatttc tttgtttttt agttttttt 3060
 aagtcaccc taacaatgaa attgtgtaga ttcaaaaata gaattaattc gtaataaaaaa 3120
 gtcgaaaaaa attgtgtcc ctccccccat taataataat tctatcccaa aatctacaca 3180
 atgttctgt tacacttctt atgtttttt tacttctgt aaatttttt tgaacatca 3240
 tagaaaaaaac cgcacacaaa atacccatc atatgttacg tttcgttta tgacccgaat 3300
 ttttatttt tcgcacgtt gggctctca tgacgtcaaa tcatgtcat cgtaaaaaaag 3360
 ttttggaga tttttggat ttttcaatca agtggaaattt ttttccctgc 3420
 ttttgcattt tgggggttcc cccattgtt ttttgcattt ttttgcattt 3480
 tgctaaaaatc acaagtattt atgagcacga tgcaagaaat atcggaaaga ggttgggtt 3540
 tgaggcttag tggaaaggta gtagaaattt ataatttggaa agtggagtag ttttgcattt 3600
 gttttgcct taaatgacag aatacattcc caatatacca aacataactg ttttgcattt 3660
 gtcggccgtt cggccctt cgtctcgcc gtttgcattt tgacggtaa aacctctgac 3720
 acatgcgtt cccggagacg gtacacgtt gtttgcattt ggttgggtt 3780
 cccgtcagg cgcgtcagg ggttgggtt ctgttgcattt tatcgccat 3840
 cagagcagat ttttgcattt tgacccatc tgccgttga aataccgcac agatcgtaa 3900
 ggagaaaaata cccgtcagg cggccattaa ggcctcgta tacgcattt ttttgcattt 3960
 aatgtcatga taataatggt ttcttagacg tcagggttgc ctttgcattt aaatgtgc 4020
 ggaaccctta ttgtttattt ttctaaata cattcaataa ttttgcattt ttttgcattt 4080
 taaccctgtt aatgttca ataatattga aaaaggaaaga gtatgatcat ttttgcattt 4140
 cgtgtccccc ttatttttcc ttttgcattt ttttgcattt ttttgcattt 4200
 acgctggta aagtaaaaaga ttttgcattt ctttgcattt ctttgcattt 4260
 ctggatctt acagcgtaa gatccttgcg ttttgcattt ctttgcattt 4320
 atgagcattt ttttgcattt ttttgcattt ttttgcattt 4380
 gagcaactcg ttttgcattt aactatttgc ttttgcattt 4440
 acagaaaaatc atcttacgaa ttttgcattt 4500
 atgagtata acactcgcc caacttactt ttttgcattt 4560
 accgctttt ttttgcattt 4620
 ctgatgtttt ttttgcattt 4680
 acgttgcgtt aactttaac ttttgcattt 4740
 gactggatgg aggccgat ttttgcattt 4800
 ttttgcattt ttttgcattt 4860
 ctggggccat atggtaagcc ttttgcattt 4920
 actatggat aacgaaatag acagatcgat ttttgcattt 4980
 taactgttagt accaagttt ttttgcattt 5040
 ttttgcattt ttttgcattt 5100
 gagttttttt ttttgcattt 5160
 cttttttttc ttttgcattt 5220
 gttttttttc ttttgcattt 5280
 ggcgttgcgtt ttttgcattt 5340
 ttttgcattt ttttgcattt 5400
 ggcgttgcgtt ttttgcattt 5460
 cttttttttc ttttgcattt 5520
 gaaactggat acctacagcg ttttgcattt 5580
 gcccgttgcgtt ttttgcattt 5640
 gggggaaacg cttttttttc ttttgcattt 5700
 cttttttttc ttttgcattt 5760
 ttttgcattt ttttgcattt 5820
 cttttttttc ttttgcattt 5880
 cttttttttc ttttgcattt 5940
 cccgttgcgtt ttttgcattt 6000
 ttttgcattt ttttgcattt 6060
 cttttttttc ttttgcattt 6120
 ttttgcattt ttttgcattt 6180

tactaactaa ctattctcat taaaattttc agagcttaaa aatggctgaa atcactcaca 6240
 acgatggata cgctaacaac ttggaaatga aataagctt catgcctgca gagaaaaaaa 6300
 atactgttt tccttgcaaa attcggtgt ttcttcaaag agaaactttt gaagtccggcg 6360
 cgagcattc cttcttgac ttctctttt ccggccaaaaaa gccttagcatt ttatgtata 6420
 atttgattac acacactcg agttcttgc catgataaaag tgtttcatgg gcaactcgccc 6480
 taacagtaca tgacaaggc ggattattatc catgcataat tgaagacaaa ctccaaatgt 6540
 gtgtcattt tggagccccg tggggggcag ctgtctcaat tatattacta gggagacgag 6600
 gagggggacc ttatcgaacg tcgcattgac catttttctt tctttatgc ctctttcac 6660
 tctctcacac attaatcgat tcataactc ccattttctt tcatgaaagggt gtgggttttt 6720
 agctttttt cccgattttt aaaaaggaga ggctgacgat gtttagaaaaa agagaacgga 6780
 gccgaaaaaaa catccgtatc aagtcttctt ttaagccga cacttttttag acagcattcg 6840
 ccgtatgtt tgaagttaa attttaaaaa ataaaaattt gtttcaattt ttttttattt 6900
 ctaaataggc aaaaggtttt tcaagaactc tagaaaaactt agcttaattt atgggtacta 6960
 gaaaaatttct tttttttaat ttaatattt tcttaagatg taattacag aagctttttt 7020
 gaaaatttctt aattttaaaaa aatttgcgtat tttagaaaaa agtcttcaga aatgatgtaaa 7080
 agctcaattt agaaggttgt tttttttttt aaaaacacgaa aaaaagacac tatttatctt 7140
 ttccctcccg cgtaaaaaattt gtttgggtga taatagtgtat ccgtgtctt tttgcactcg 7200
 gctttcaca ccgtgtttcc tcttcaatttga cccaaacagga aaaaaaaaca tcacgtctga 7260
 gacgggtgaat tgccttattca agagcgtcgt ctcttttacc cagtaacaaa aaaaattttgg 7320
 ttcttttact ttatattttt gtaggtcaca aaaaaaaagt gatgcagggt tgggggtcg 7380
 ttgtctccac accaccccg cctccagcag cacacatca ttttcgtgt ttttcgacga 7440
 ttcttggat gccgcggcgtc tgaatgcacc cgcacacta caccacactc 7500
 actttcggtt gtattactac acgtcatcgat ttgtcgatg ctcccgctct ttgttcccca 7560
 ctcaacttcc attttttttt ttgttgcattt gatttttttt aaatggtaca ccactcctga 7620
 cgtttctacc ttcttgcattt ccgtccattt agattttata tggaaattttt tttttttttt 7680
 tagggccagag agttcttagt ttgttctaa aagtcttagt cagacatca ttttcttattt 7740
 ctcatcaaaa aaaaagggttga taaagaaaaac ttgttattca gaaagagtgt gtctcggtga 7800
 aatttgcattca aaaaaaaatcc cccacccctt gtttgcattt caaaatgtt gatcaacgga 7860
 ttttttccctt ctgcattcaat ttgttgcattt cgtctgtctt gccaaagggtt gtgttccga 7920
 gaaaaagatg agagaatttta caacagaaaaa tggaaaaaaat tggcccaat aatgaatgtt 7980
 tatccgagat tgatggaaaa gatattaatg ttctttacgg tttggagggg agagagagat 8040
 agattttcgc atcaactcc gccttttaca tttttttttt aatctttttt aatctttttt 8100
 catcatttt aatagaaaaat cgagaaatttca cagtaatttca gcaattttctt tgccaaaaat 8160
 acacgaaaaattt tttttttttt gtttgcattt gtttgcattt gtttgcattt gtttgcattt 8220
 ttataaaattt tcaaaatttca gtttgcattt tccgcataat tggacatcaat tttttttttt 8280
 gtcatttttca tcaacaagaa atcgtaaaa ttcttgcattt tttttttttt tttttttttt 8340
 atacacgaaaaa tatatgttaa ttcccgaaaaa tatttgcattt tccgcacca gtttgcattt 8400
 agtggccagg attttatcgat aaaaagggtt gcttgcattt cttttttttt tttttttttt 8460
 cttaattttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 8520
 attttcgtcg tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 8580
 accatgatat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 8640
 ggttgcatttcc accgcgcattt tttttttttt tttttttttt tttttttttt tttttttttt 8700
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 8760
 agccgagaaat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 8820
 tggccgcgc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 8880
 tgccacccatca tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 8940
 ttgtcgatca tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 9000
 gcttgcatttcc ggttgcattt tttttttttt tttttttttt tttttttttt tttttttttt 9060
 ttcttcattttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 9120
 tagcattttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 9180
 ttgggttaca gccgtcaaca gatccccggg attggccaaa ggacccaaag gtatgttgc 9240
 aatgatacta acataacata gacatattttc aggaggaccc ttgttgcattt ggttccgatc 9300
 tcagaaaaaa 9309